AD_____

Award Number: W81XWH-07-1-0545

TITLE: Highly conserved, large, non-coding transcripts and their role in the development of breast cancer

PRINCIPAL INVESTIGATOR: David I Smith, Ph.D.

CONTRACTING ORGANIZATION: Mayo Clinic
Rochester, MN 55905

REPORT DATE: August 2008

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
31-08-2008	Annual	1 Aug 2007-31 Jul 2008

4. TITLE AND SUBTITLE
Highly conserved, large, non-coding transcripts and their role in the development of breast cancer

5a. CONTRACT NUMBER
W81XWH-07-1-0545

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
David I Smith, Ph.D.

Email: smith.david@mayo.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Mayo Clinic
200 1st Street SW
Rochester, MN 55905

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research &
Materiel Command
Fort Detrick, Maryland

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The sequencing of the human genome has revealed that only 2% of the genome actually codes for protein. However, the remainder of the genome is not "junk" and it has recently been revealed that most of the genome is transcriptionally active. We utilized a tiling array approach to examine the entire genome for transcriptional activity and found a large number of non-coding transcripts. When we originally proposed the work in this Concept Award, we were proposing to study a group of long, highly conserved, non-coding transcripts which had altered expression and sometimes mutations in breast cancer. We have subsequently found that many of these highly conserved transcripts are either part of known genes or highly homologous to known genes. However, we've now identified two new groups of novel non-coding transcripts which are not part of genes. The first group are non-coding transcripts that have increased expression in response to the DNA damage induced by the carcinogen NNK. The second group were identified by analyzing breast cancer cell lines with tiling arrays searching for non-coding transcripts that had consistently altered expression. We have requested and obtained a no-cost extension for this work and will begin to characterize these transcripts exactly as proposed in our original Concept Award to determine the role they play in the development of breast cancer.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	11	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

ABSTRACT

The sequencing of the human genome has revealed that only 2% of the genome actually codes for protein. However, the remainder of the genome is not “junk” and it has recently been revealed that most of the genome is transcriptionally active. We utilized a tiling array approach to examine the entire genome for transcriptional activity and found a large number of non-coding transcripts. When we originally proposed the work in this Concept Award, we were proposing to study a group of long, highly conserved, non-coding transcripts which had altered expression and sometimes mutations in breast cancer. We have subsequently found that many of these highly conserved transcripts are either part of known genes or highly homologous to known genes. However, we’ve now identified two new groups of novel non-coding transcripts which are not part of genes. The first group are non-coding transcripts that have increased expression in response to the DNA damage induced by the carcinogen NNK. These NNK-induced transcripts (NITs) are all over 300 nucleotides long and have altered expression in breast cancer. We have validated these transcripts and have utilized Northern blots to determine the precise size of these transcripts (and they are between 500 and 1,500 base pairs in length). The second group were identified by analyzing breast cancer cell lines with tiling arrays searching for non-coding transcripts that had consistently altered expression. We’ve now identified a group of breast cancer non-coding transcripts (bcNCTs). These have been validated in a larger panel of breast cancer cell lines and the precise size of these transcripts have been determined using Northern blot analysis. We have requested and obtained a no-cost extension for this work and will begin to characterize these transcripts exactly as proposed in our original Concept Award to determine the role they play in the development of breast cancer.

INTRODUCTION

BODY

Original Proposal

We had identified a group of putative long highly conserved non-coding transcripts that we proposed to characterize further. There were three Specific Aims to our proposal. They were:

Specific Aim #1: Characterization of NCT4 and NCT 5 (the two most highly conserved non-coding transcripts that we had identified) in the normal breast epithelial cell lines HMEC and MCF10

Specific Aim #2: Determining if NCT4 and/or NCT5 were mutational targets in a panel of breast cancer cell lines

Specific Aim #3: To utilize siRNA technology to abrogate the expression of NCT4 and NCT5 in normal breast cell lines and then to determine what effect this has on these cells and on the expression of all the coding transcripts

Only 2% of the human genome actually codes for protein. In spite of this, it turns out that most of the genome is transcriptionally active. The question we set out to answer is what the function of these non-coding transcripts are and what role, if any, do they play in the development of breast cancer. Our major hypothesis is that these non-coding transcripts play an important regulatory role within the cells and would be important targets of alteration during the development of breast cancer.

We decided to focus our efforts on a group of long (greater than 400 nts) non-coding transcripts that we had identified by using a tiling array approach to identify transcriptional units across the genome. We specifically focused on the most highly conserved on these non-coding transcripts and had preliminary evidence that several of these were also targets of alteration (both in terms of expression and as occasional mutational targets) in breast cancer. We proposed to characterize two of these (NCT4 and NCT5) as they were the most highly conserved.

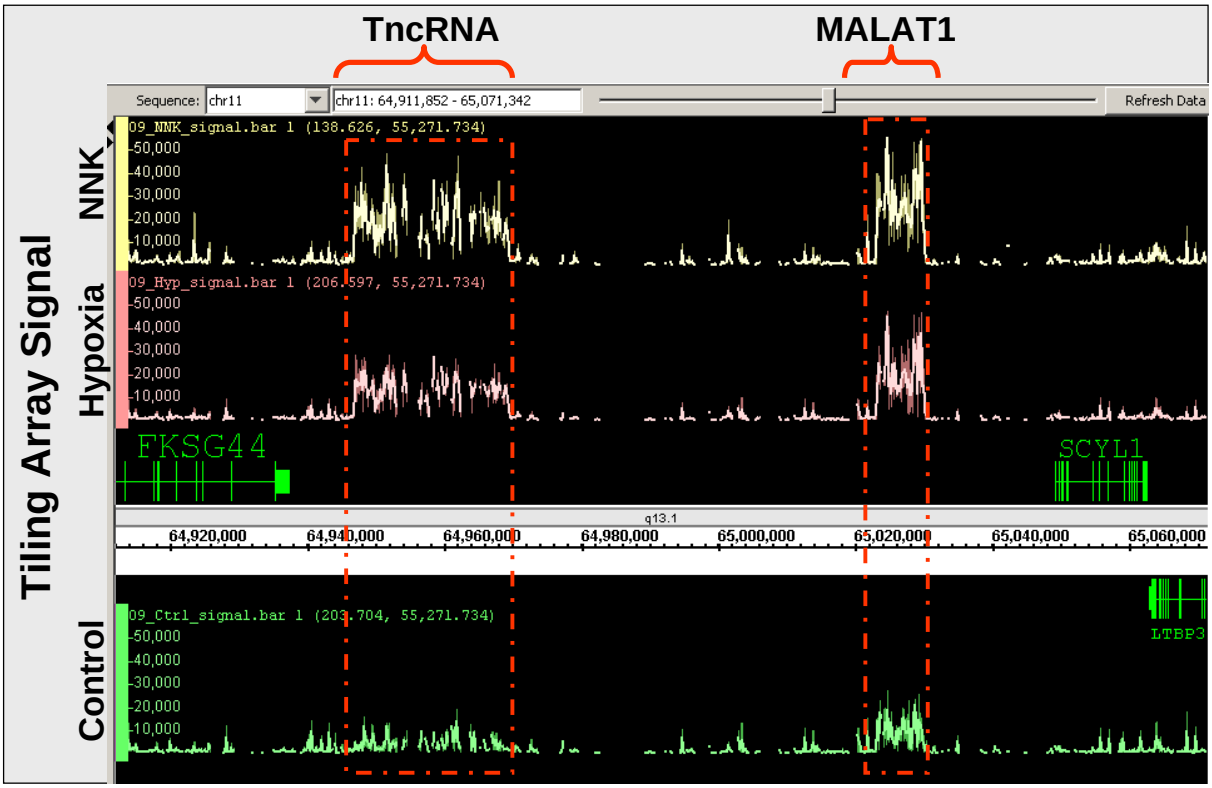
However, subsequent analysis of the highly conserved non-coding transcripts has revealed that most of the more highly conserved transcripts actually had homology to existing coding transcripts. Indeed results of the ENCODE project now reveal that each coding gene produces on the average about 5 distinct transcripts and that there is much greater complexity to gene organization than previously anticipated. In addition, the simple model of genes being merely a collection of contiguous exons that are spliced together may also be wrong. Many transcripts generated are actually produced from quite disparate chromosomal regions (this was revealed most convincingly by doing mate-pair sequence analysis of large numbers of transcripts using Next Generation DNA sequencing technology). In addition, there are cryptic exons that are sometimes hundreds of kilobases upstream or downstream from the simple organized gene. Indeed, many of the highly conserved non-coding transcripts that we were characterizing were found to be linked to known genes either due to extensive homology to those known genes, or because they actually corresponded to those distant exons.

Fortunately, we had two additional sources of non-coding transcripts that we began to analyze in greater detail. When we performed our initial tiling array experiment to identify possible non-coding transcripts we not only examined a normal epithelial cell line but also exposed that cell line to two types of stresses that have been associated with the development of cancer. The first was exposure to the DNA damaging carcinogen NNK. The second was growth under hypoxic conditions.

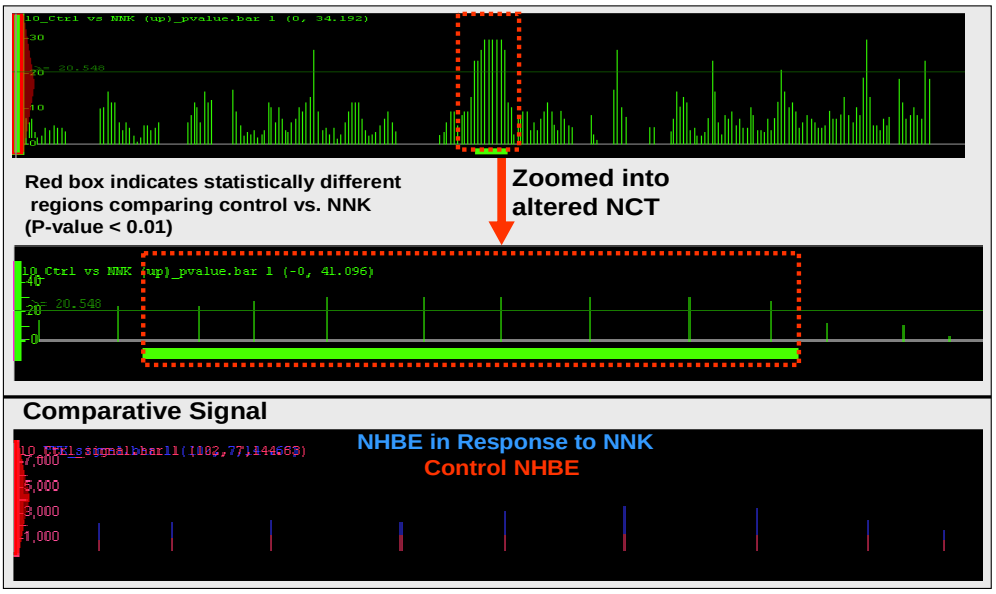
We analyzed the entire genome in response to these two different types of stress looking for potential non-coding transcripts that were either induced or repressed by one, or both, of these stresses. This analysis identified transcripts that were induced by exposure to NNK (called NITs for NNK-induced transcripts), suppressed by exposure to NNK (called NSTs for NNK-suppressed transcripts), induced by exposure to hypoxia (called HITs for hypoxia-induced transcripts) and finally suppressed by exposure to hypoxia (called HSTs for hypoxia-suppressed transcripts).

We continued to focus our attention on longer transcripts as we were attempting to look for novel transcripts that were not either miRNAs or potential miRNA precursors. We did not look for conserved non-coding transcripts, as our previous studies had shown that most of these were actually either homologous to or part of existing known coding genes. It is interesting and important to note that two previously reported non-coding transcripts tncRNA and MALAT-1 (which are located adjacent to each other on chromosome 11) were identified in this screen as being induced by exposure to NNK, hence qualify as NITs. The Figures below show the integrated genome browser (IGB) results with tncRNA and MALAT-1, as well as several of the newly identified stress-responsive non-coding transcripts.

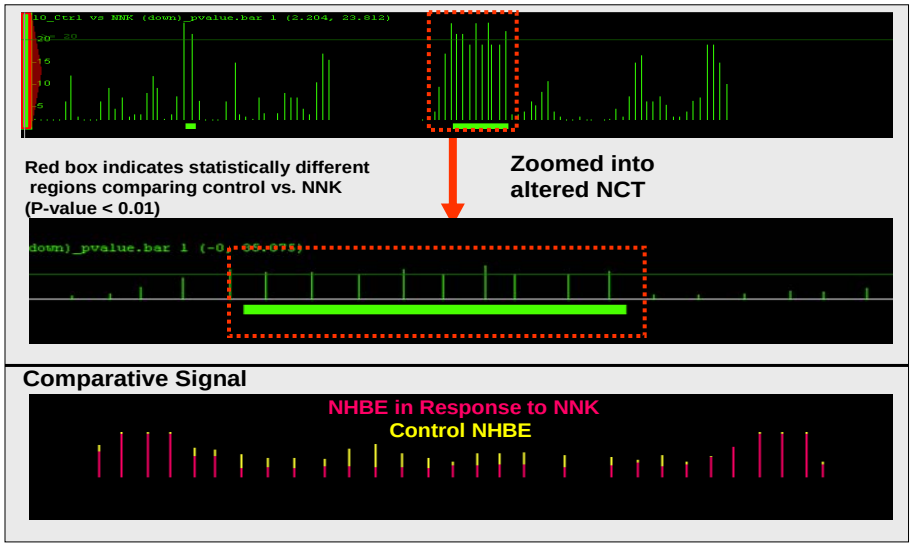
Stress Response of Long ncRNAs



NNK-Induced NCT (NITs)

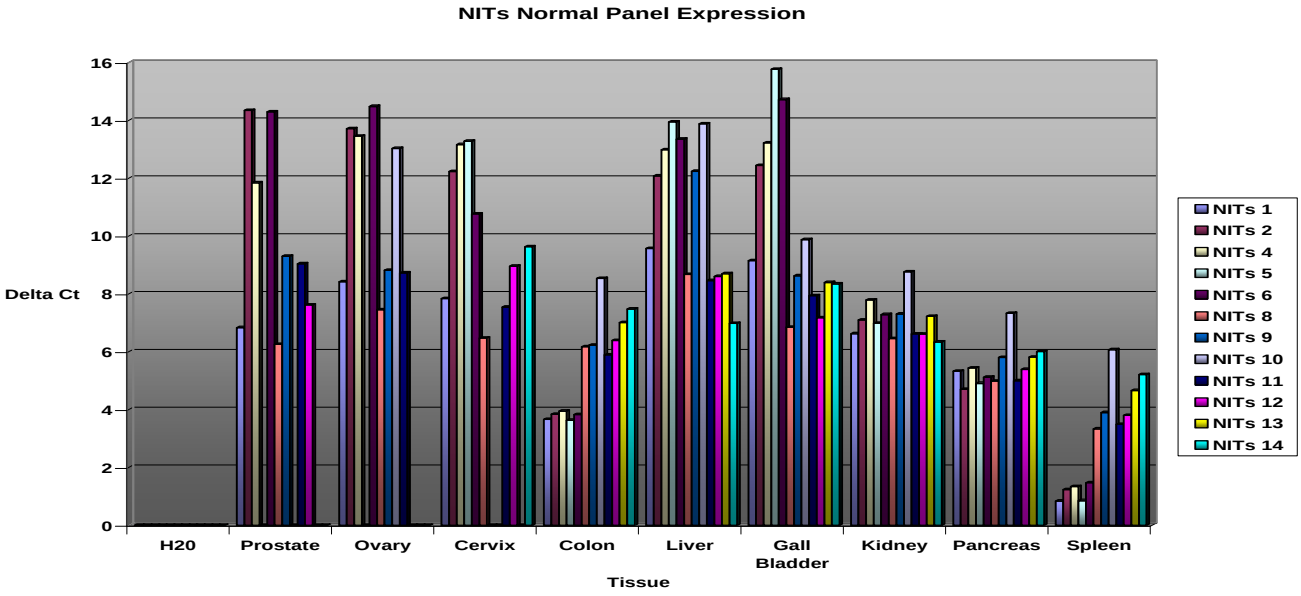


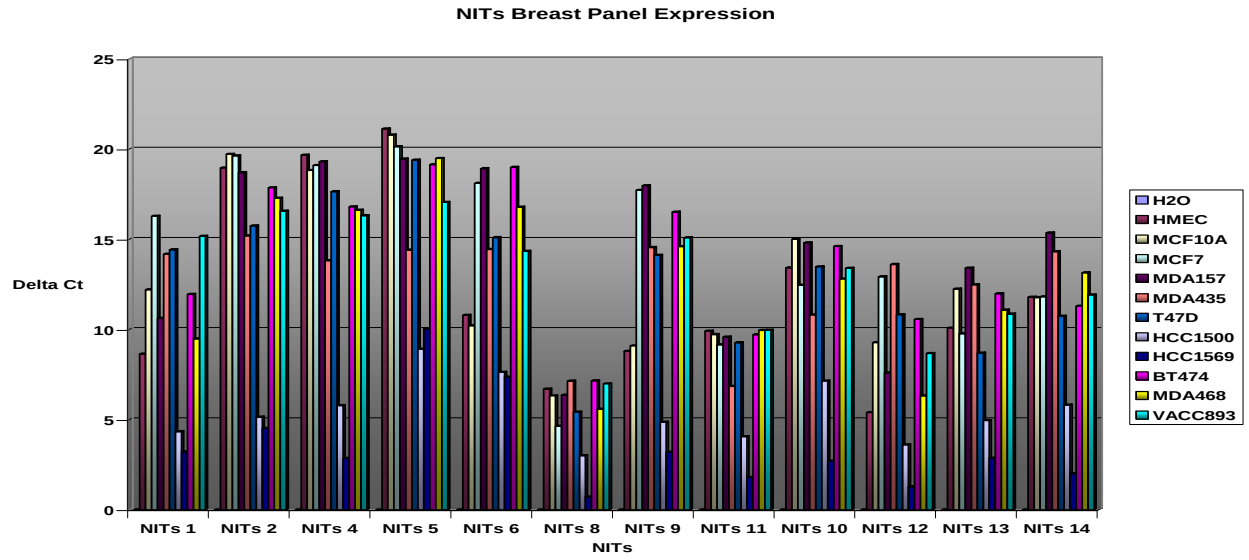
NNK-Suppressed NCT (NSTs)

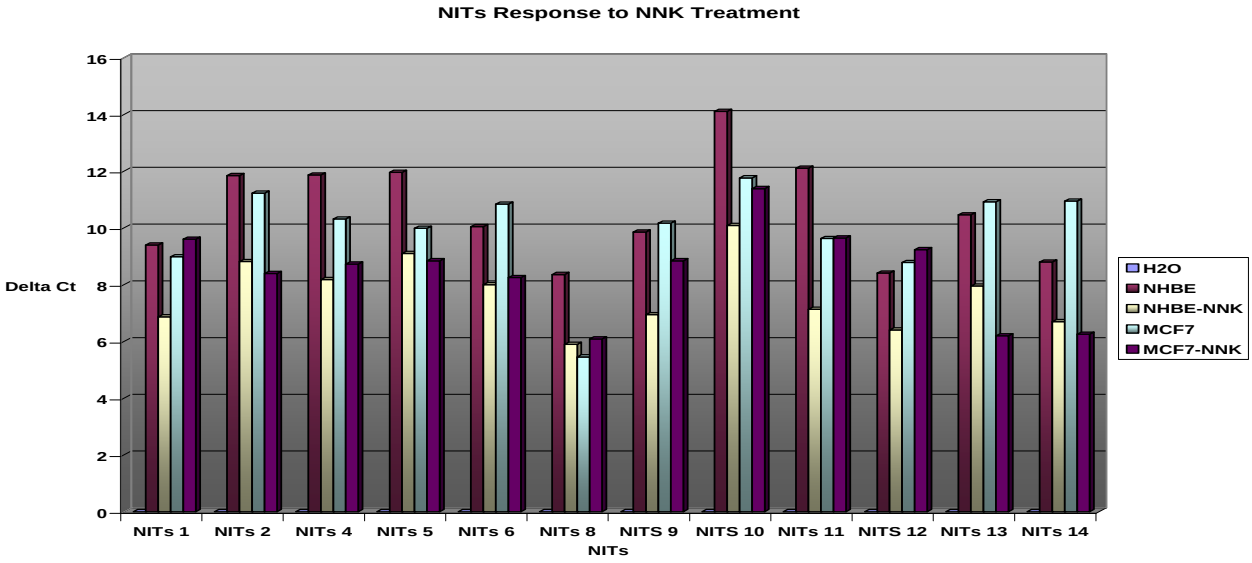


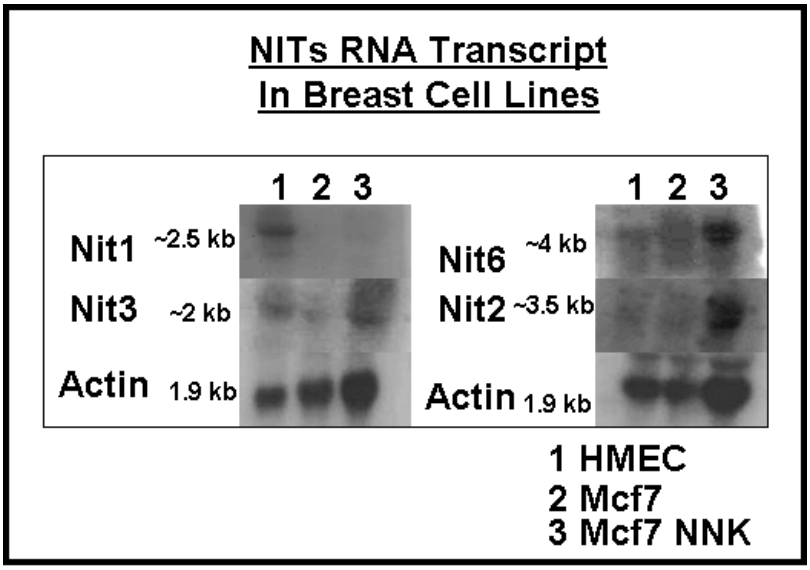
We next have begun to more fully characterize the NITs and the NSTs (we have not yet begun to work on the HITs and HSTs). We constructed real-time PCR primers to amplify each of the 14 NITs and 22 NSTs and determined optimal concentrations of the primers for real-time PCR. Once this was accomplished we began to analyze panels of random-primer primed cDNAs made from different RNA samples. This included a panel of RNAs isolated from various normal human tissues, a panel of breast normal and cancer-derived cell lines, as well as various normal cell lines exposed to either NNK or growth under hypoxic conditions. The goal of this was to determine whether these transcripts were indeed stress responsive (as we had observed in the tiling array experiment, hence validation of those results), the spectrum of their expression in different tissues, and finally whether or not they have altered expression in both breast cancer cell lines and primary tumors.

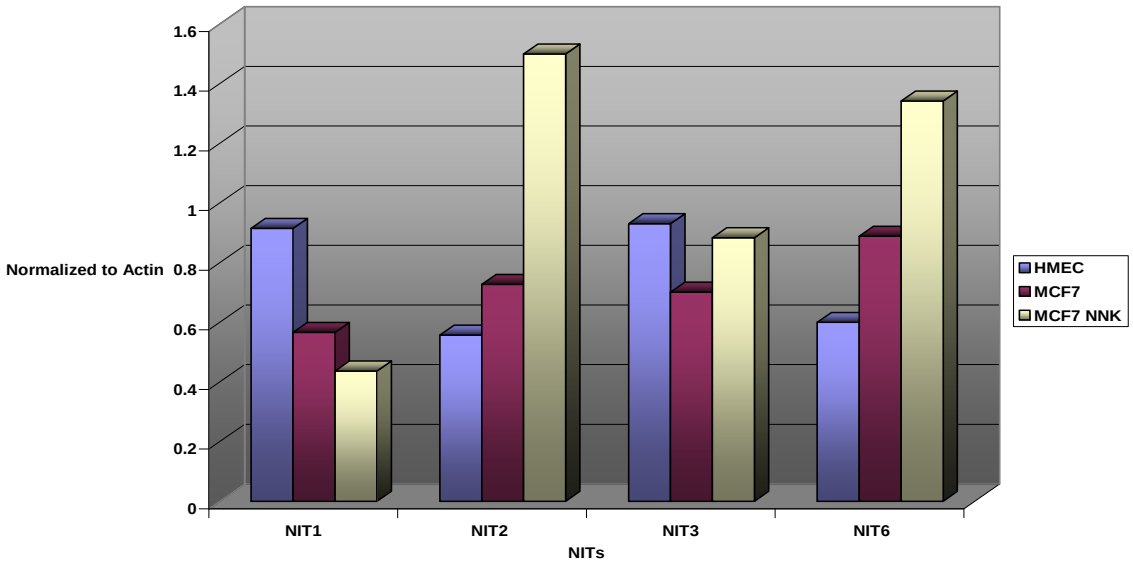
We also generated probes for these stress-responsive non-coding transcripts and hybridized them to Northern Blots to determine the size of the putative non-coding transcripts. The figure below shows several representative Northern blots. This analysis revealed that the full size of these transcripts was greater than that determined on the tiling arrays.




NITs Breast Panel Expression
Delta Ct
25
20
15
10
5
0
NITs 1
NITs 2
NITs 4
NITs 5
NITs 6
NITs 8
NITs 9
NITs 11
NITs 10
NITs 12
NITs 13
NITs 14
NITs
H2O
HMEC
MCF10A
MCF7
MDA157
MDA435
T47D
HCC1500
HCC1569
BT474
MDA468
VACC893

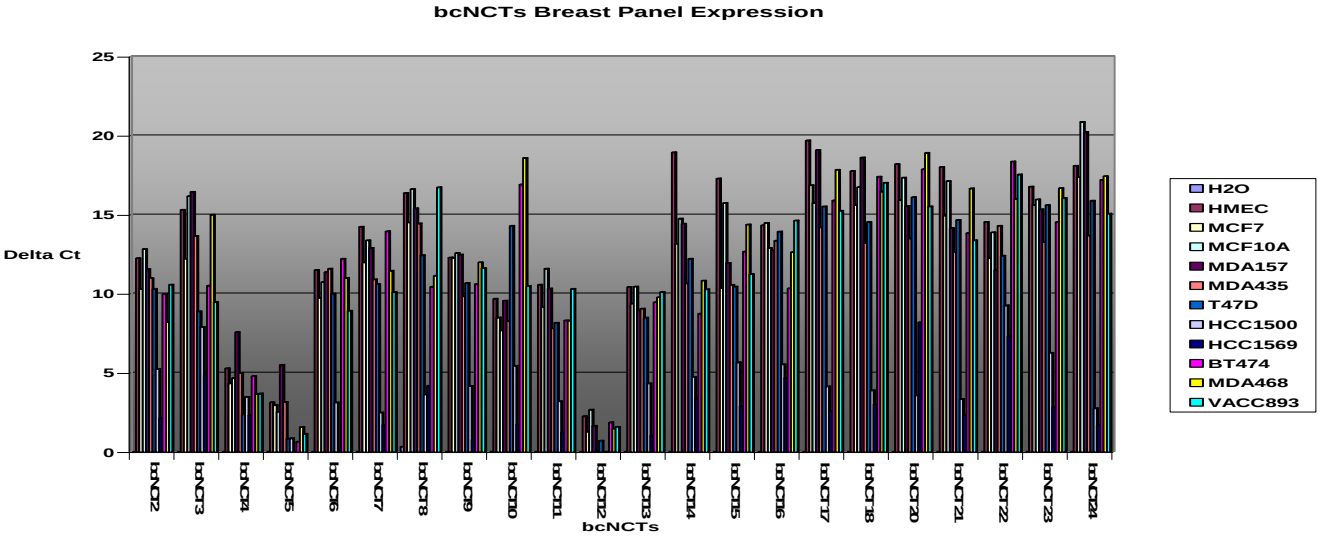

NITs Response to NNK Treatment
Delta Ct
16
14
12
10
8
6
4
2
0
NITs 1
NITs 2
NITs 4
NITs 5
NITs 6
NITs 8
NITs 9
NITS 10
NITs 11
NITS 12
NITs 13
NITs 14
NITs
H2O
NHBE
NHBE-NNK
MCF7
MCF7-NNK

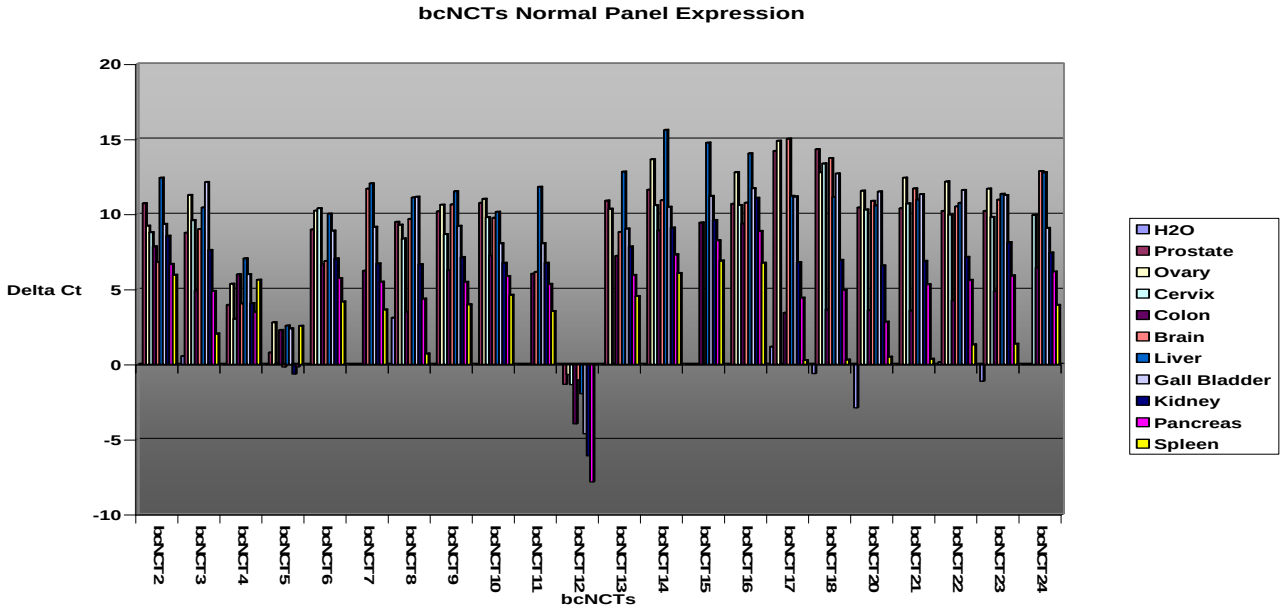

NITs RNA Transcript
In Breast Cell Lines
1 2 3
Nit1 ~2.5 kb
Nit3 ~2 kb
Actin 1.9 kb
1 2 3
Nit6 ~4 kb
Nit2 ~3.5 kb
Actin 1.9 kb
1 HMEC
2 Mcf7
3 Mcf7 NNK

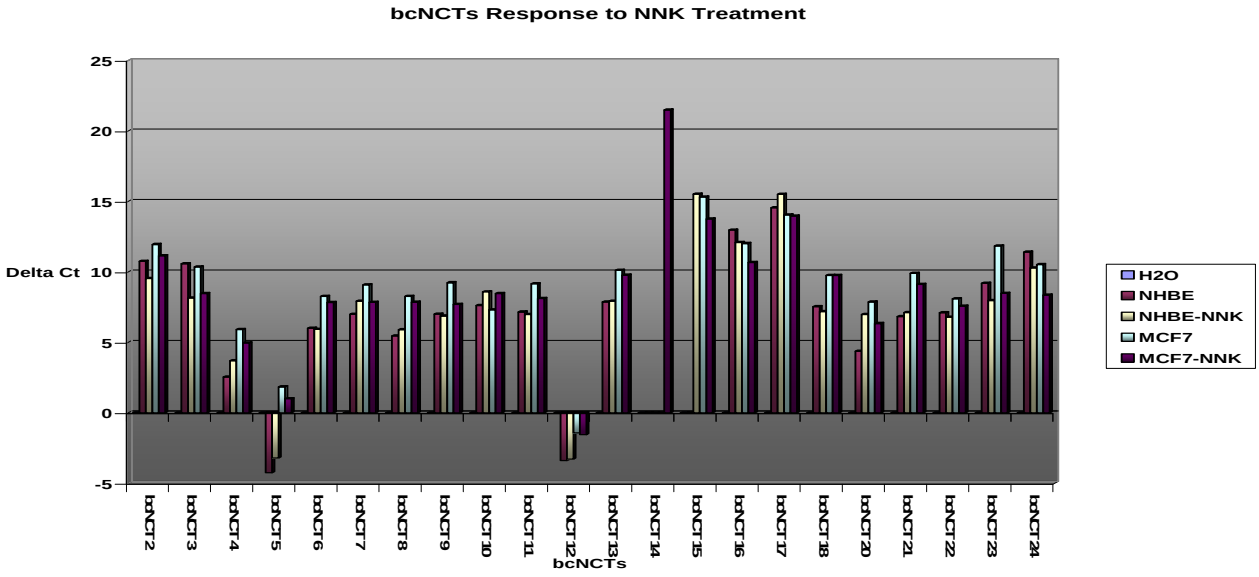


The second source of potential non-coding transcripts worth studying were derived from an experiment that we conducted to directly search for non-coding transcripts that were altered in breast cancer cell lines. We took the normal breast epithelial cell line HMEC and four breast cancer cell lines including HCC 1500, HCC1569, MDA157, T47D, and MDA 435. Total RNA was isolated from each of these and then hybridized to a single tiling array chip in triplicate. While our initial studies were done with 14 tiling array chips covering the entire genome we decided to conduct a more focused (and considerably less expensive) experiment of only examining a single tiling array chip which had probes covered three human chromosomes (chromosomes 8, 11, and 12). We then searched for non-coding transcripts which were altered in more than one breast cancer cell line, relative to HMEC. While there were many transcripts altered in just one of the cell lines, common alterations in multiple cells lines were few. There were 24 transcripts altered in two cell lines, and 1 transcript altered in three cell lines. No transcripts were found to be altered in all four cell lines and these transcripts had no correlation with NNK responsive NCTs. The transcripts that had these consistent alterations were called bcNCT (for breast cancer Non-Coding Transcripts).

The 24 bcNCTs alterted in the two cell lines were then characterized exactly as the NITs and NSTs were characterized. (The other altered transcripts around 300 are in the process of being analyzed.) Primers were derived to amplify the majority of the regions encoding them and these were used to amplify the genomic region to be utilized as probes for Northern blot hybridizations. Real-time primers were derived to verify whether or not the bcNCTs were indeed consistently altered in a much larger panel of breast cancer cell lines, as well as primary tumors. We also characterized the bcNCTs to determine expression in normal human tissues. The bcNCTs were also characterized to identify if they were stress responsive, as this was our original goal with the stress response tiling array experiment. It is important to note there that the two characterized non-coding transcripts, tncRNA and MALAT-1, which have increased expression in response to DNA damage induced by NNK also have increased expression in several of the breast cancer cell lines. This provides us good support for our hypothesis that stress-responsive non-coding transcripts are good candidates for transcripts that also have altered expression during the development of breast cancer. The Figures below show several of the bcNCTs that we have chosen for further study.


bcNCTs Breast Panel Expression
Delta Ct
25
20
15
10
5
0
bcNCTs
H2O
HMEC
MCF7
MCF10A
MDA157
MDA435
T47D
HCC1500
HCC1569
BT474
MDA468
VACC893


bcNCTs Normal Panel Expression
Delta Ct
20
15
10
5
0
-5
-10
bcNCTs
H2O
Prostate
Ovary
Cervix
Colon
Brain
Liver
Gall Bladder
Kidney
Pancreas
Spleen


bcNCTs Response to NNK Treatment
Delta Ct
25
20
15
10
5
0
-5
bcNCTs
H2O
NHBE
NHBE-NNK
MCF7
MCF7-NNK

We would like to thank the Department of Defense Breast Cancer Program for supporting our work. While it is still in the preliminary stages, we remain hopeful that this work will identify important new targets of alteration during the development of breast cancer. We have submitted an RO1 proposal to the NIH on some of the non-coding transcripts, but as expected, the first submission was triaged out. However, with additional supportive results (which we hope to obtain with Department of Defense support), this grant should be more competitive.

We have asked for, and received, a no-cost extension on our Breast Cancer Concept award. Our plans over the next several months will be to examine each of the different non-coding transcripts identified (which will include both the stress responsive non-coding transcripts as well as the bcNCTs described above) and to determine which of them are good candidates for further characterization. At that point we will take those candidates and examine them exactly as originally described for the highly conserved NCTs. Our goal will be to demonstrate some functional significance to breast cancer of one, or more, of the identified non-coding transcripts. At that point this work would be considerably more exciting and thus capable of getting support either from the NIH or from a Department of Defense Breast Cancer Idea award.

KEY RESEARCH ACCOMPLISHMENTS

(1) Identified NNK-induced long non-coding transcripts (NITs).
(2) Validated these NIT transcripts both with real-time RT-PCR and with Northern blot analysis.
(3) Identified breast cancer altered long non-coding transcripts (bcNCTs).
(4) Validated these bcNCT transcripts both with real-time RT-PCR and with Northern blots.

PUBLICATIONS AS A DIRECT RESULT OF THIS GRANT

Perez DS, Hoage TR, Pritchett JR, Ducharme-Smith AL, Halling ML, Ganapathiraju SC, Streng PS, Smith DI. Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Molec Genet* 2008; 17: 642-655.

Silva J, Perez DS, Ducharme-Smith AL, Pritchett JR, Smith DI. NNK-induced transcripts: A new group of large stress responsive non-coding transcripts have altered expression in breast cancers. Manuscript in preparation.